K-Nearest Neighbors (K-NN) Regression

Question

You are given a dataset with two input features x_1 and x_2, and a continuous output y. The dataset is as follows:

S. No.	x_1	x_2	y
1	4	1	2.5
2	2	4	3.0
3	2	3	1.8
4	3	6	2.0
5	4	4	4.5
6	9	10	7.0
7	6	8	6.5
8	9	5	5.8
9	8	7	6.2
10	10	8	7.5

Using the K-Nearest Neighbors (K-NN) regression algorithm with $k = 3$, predict the output y for a new data point $(x_1, x_2) = (5, 8)$. Show all mathematical steps and calculations leading to the predicted output.

Solution

Step 1: Calculate Euclidean Distances

The Euclidean distance between two points (x_1, x_2) and (x'_1, x'_2) is given by:

$$d = \sqrt{(x_1 - x'_1)^2 + (x_2 - x'_2)^2}$$

We compute the Euclidean distance between the point $(5, 8)$ and all other points in the dataset.

S. No.	(x_1, x_2)	y	Distance from $(5, 8)$
1	$(4, 1)$	2.5	$\sqrt{(5-4)^2 + (8-1)^2} = \sqrt{1+49} = \sqrt{50} \approx 7.07$
2	$(2, 4)$	3.0	$\sqrt{(5-2)^2 + (8-4)^2} = \sqrt{9+16} = \sqrt{25} = 5.00$
3	$(2, 3)$	1.8	$\sqrt{(5-2)^2 + (8-3)^2} = \sqrt{9+25} = \sqrt{34} \approx 5.83$
4	$(3, 6)$	2.0	$\sqrt{(5-3)^2 + (8-6)^2} = \sqrt{4+4} = \sqrt{8} \approx 2.83$
5	$(4, 4)$	4.5	$\sqrt{(5-4)^2 + (8-4)^2} = \sqrt{1+16} = \sqrt{17} \approx 4.12$
6	$(9, 10)$	7.0	$\sqrt{(5-9)^2 + (8-10)^2} = \sqrt{16+4} = \sqrt{20} \approx 4.47$
7	$(6, 8)$	6.5	$\sqrt{(5-6)^2 + (8-8)^2} = \sqrt{1+0} = \sqrt{1} = 1.00$
8	$(9, 5)$	5.8	$\sqrt{(5-9)^2 + (8-5)^2} = \sqrt{16+9} = \sqrt{25} = 5.00$
9	$(8, 7)$	6.2	$\sqrt{(5-8)^2 + (8-7)^2} = \sqrt{9+1} = \sqrt{10} \approx 3.16$
10	$(10, 8)$	7.5	$\sqrt{(5-10)^2 + (8-8)^2} = \sqrt{25+0} = \sqrt{25} = 5.00$

Step 2: Select the 3 Nearest Neighbors

We sort the distances in ascending order:

S. No.	Distance	y
7	1.00	6.5
4	2.83	2.0
9	3.16	6.2
5	4.12	4.5
6	4.47	7.0
2	5.00	3.0
8	5.00	5.8
10	5.00	7.5
3	5.83	1.8
1	7.07	2.5

The 3 nearest neighbors are:

S. No.	Distance	y
7	1.00	6.5
4	2.83	2.0
9	3.16	6.2

Step 3: Predict the Output (Average the y values)

We predict the output by taking the average of the y values of the three nearest neighbors:

$$\text{Predicted } y = \frac{6.5 + 2.0 + 6.2}{3} = \frac{14.7}{3} = 4.9$$

Thus, the predicted continuous output for the point $(5, 8)$ is $\boxed{4.9}$.